

The origin of fat tailed distributions in financial time series

G. M. Viswanathan¹, U. L. Fulco^{1,2}, M. L. Lyra¹ and M. Serva^{1,3}

¹ *Departamento de Física, Universidade Federal de Alagoas,
Maceió-AL, 57072-970, Brazil*

³ *Departamento de Física, Universidade Federal de Piauí,
Teresina-PI, 64049-550 , Brazil and*

³ *Dipartimento di Matematica and I.N.F.M. Università dell'Aquila, I-67100 L'Aquila, Italy*

(Revised 06 November 2002)

Abstract

A classic problem in physics is the origin of fat tailed distributions generated by complex systems. We study the distributions of stock returns measured over different time lags τ . We find that destroying all correlations without changing the $\tau = 1$ d distribution, by shuffling the order of the daily returns, causes the fat tails almost to vanish for $\tau > 1$ d. We argue that the fat tails are caused by known long-range volatility correlations. Indeed, destroying only sign correlations, by shuffling the order of only the signs (but not the absolute values) of the daily returns, allows the fat tails to persist for $\tau > 1$ d.

PACS numbers: 89.65.Gh, 05.40.-a, 05.40.Fb

Over the last few decades, remarkable progress has been made in quantitatively describing non-Gaussian phenomena, including those observed in economic [1] and social [2] systems, that are typically characterized by the presence of fat tailed Lévy distributions [3]. The behavior of financial markets has recently [1, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13] become a focus of interest to physicists as well as an area of active research because of its rich and complex dynamics [14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29]. One open question relates to the probability distribution underlying returns on stock markets. It is well known that the century-old Gaussian model [30] underestimates the probability of large events. Indeed, the distribution of stock returns is fat tailed [1, 13, 14]. On the one hand, the fat tails could be due to an underlying Lévy distribution. According to the generalized central limit theorem, the sum of τ independent (i.e., uncorrelated) Lévy distributed random variables is also Lévy distributed, such that the persistence of fat tails for large τ is due solely to the Lévy nature of the original ($\tau = 1$) distribution. Furthermore, if the $\tau = 1$ distribution is Lévy with exponential truncation of the tails, then we still expect a certain degree of persistence of the fat tails [31]. On the other hand, fat tails can also persist [10] for $\tau > 1$ due to long-range correlations in a “hidden variable” [8] such as volatility (i.e., locally measured standard deviation) [6]. Moreover, such long-range correlations have been found to produce fat tails [28]. Although stock returns lack long-range power law correlations, yet the absolute values of the returns are known to be long-range correlated [8, 9, 11, 32, 33, 34, 35]. The absolute returns are power law correlated with non-unique scaling exponents [8, 11]. Here we test the hypothesis (see ref. [10]) that long-range volatility correlations are the origin of the fat tails.

Below we will show that for stock market returns the observed persistence of fat tails for large τ cannot be explained without long-range correlations in the volatility. Shuffling the daily returns has the effect of destroying all correlations while maintaining unchanged the $\tau = 1$ d distribution. For shuffled data, we will show that the distribution is fat tailed for $\tau = 1$ d but not for $\tau > 1$ d. We interpret this finding as evidence that volatility correlations rather than the $\tau = 1$ d Lévy-like distribution are responsible for the existence of fat tails for large τ . Indeed, we will also show that shuffling only the signs of the returns allows the fat tails to persist for $\tau > 1$ d and the distribution does not converge to a Gaussian. These findings conclusively prove that known long-range volatility correlations (rather than known short-range [12] sign correlations) are responsible for fat tails for any $\tau > 1$ d. We will also

show that, remarkably, the short-range (1–2 d) sign correlations also play an important role in the distribution properties of small price changes, and we will propose an explanation for why a Gaussian fits the data so well for small, but not large, returns.

Our dataset consists of the base 10 logarithms of daily returns obtained from 59 stock market indices (obtained from yahoo.com: AEX, AORD, ATG, ATX, BFX, BSESN, BVL30, BVSP, CCSI, DJA, DJI, DJT, DJU, DOT, FCHI, FTSE, HEX, HSI, IBC, IGRA, IIX, IPSA, IXIC, JKSE, KFX, KLSE, KS11, KSE, MERV, MID, MTMS, MXX, N225, NDX, NTOT, NYA, NZ40, OEX, PSE, PSI, PX50, RUA, RUI, RUT, SAX, SETI, SML, SMSI, SOOX, SPC, SSEC, SSMI, STI, TA100, TSE, TWII, VLIC, XMI, XU100). The returns $r(t)$ are defined in terms of the prices $P(t)$ by

$$r(t) \equiv \log_{10} \frac{P(t)}{P(t-1)} . \quad (1)$$

We normalize the returns to unit variance for each market index separately. To be able to compare returns measured over differing time scales, we also define a rescaled return r_τ by

$$r_\tau(t) \equiv \frac{1}{\sqrt{\tau}} \sum_{t'=t-\tau+1}^t r(t') , \quad (2)$$

with τ measured in days and $r_1 = r$. The daily and rescaled returns play roles similar to those played by “bare” and “dressed” quantities in field theory. Note that for uncorrelated (independent) and unitary Gaussian distributed returns, their variances will be identical due to the central limit theorem: $\sigma(r_1) = \sigma(r_\tau) = 1$ d. Similarly, if $r_1(t)$ are Lévy distributed, then $r_\tau(t)$ will also be Lévy distributed.

Even for Gaussian returns, however, the presence of correlations can lead to anomalous behavior, such that r_1 and r_τ may have non-identical probability distributions. We therefore develop a method to “subtract” the effects of correlations. For each of the 59 time series, we generate a modified control time series by shuffling the order of the daily returns (Fig. 1). This shuffled daily returns model (SDRM) will have a probability distribution identical to the real data for $\tau = 1$ d, but lacks all correlations. Hence, for $\tau > 1$ d the real data and the SDRM will in general not have identical distributions (Fig. 2) unless correlations are lacking. Thus, we now have a way to test the hypothesis that the fat tails in $p(r_\tau)$ persist solely due to correlations. If the probability density distribution $p(r_\tau)$ is fat tailed for the real data but not for the SDRM, then the conclusion would be that the fat tails in $p(r_\tau)$ are due to correlations.

Shuffling the data destroys all kinds of correlations—indeed, the data become independent numbers. Specifically, shuffling destroys the known long-range power law correlations in the volatility of the returns as well as the short-range correlations in the signs of the returns. We thus develop a method for “subtracting” only the correlations in the signs of the returns $r_1(t)$ while preserving (volatility) correlations in the absolute returns $|r_1(t)|$ (Fig. 1). For each of the 59 time series, we generate a second modified control time series by shuffling the order of the signs—but not of the absolute values—of the daily returns. This shuffled signs return model (SSRM) will have a symmetrized probability distribution identical to the real data and to the SDRM for $\tau = 1$ d, but not necessarily for $\tau > 1$ d.

We study the symmetrized probability density distribution function $p(r_\tau)$ of the returns r_τ from 59 stock markets and compare them to those of the SDRM and the SSRM. We focus on the fat tailed regions of the distributions by studying a properly defined modified characteristic function

$$\begin{aligned} f(\tau) &= \int dr_\tau p(r_\tau) \exp[-(|r_\tau| - r_0)^2] \\ &\simeq (1/N) \sum_t \exp[-(|r_\tau| - r_0)^2] , \end{aligned} \quad (3)$$

where t is time in days. In order to study the fat tailed region while retaining good statistics, we chose a value $r_0 = 5$ corresponding to 5 standard deviations. (We also studied higher moments, but these are extremely sensitive to large events, rendering the results not statistically significant.) Similarly, to study the central bell curve region, we define a second function

$$g(\tau) = \int dr_\tau p(r_\tau) \exp(-r_\tau^2) \simeq (1/N) \sum_t \exp(-r_\tau^2) , \quad (4)$$

where $N = \sum_t 1$. In practice, we calculate these functions directly from the returns, rather than through the distributions, to obtain better statistics.

We find that the fat tails almost disappear for $\tau > 1$ d in the SDRM, showing that correlations are necessary for maintaining the fat tails for $\tau > 1$. This finding is consistent with the results reported in ref. [10] and rules out the possibility that the daily Lévy-like distribution is responsible for the persistence of fat tails. In fact, if this were so, the fat tails would persist for $\tau > 1$ d even after shuffling the order of the returns $r_1(t)$, contrary to our findings. One must conclude that the fat tails are mainly due to correlations. Note, however, that for the SDRM, the fat tails do not disappear entirely and $p(r_\tau)$ never becomes truly

Gaussian even for $\tau \rightarrow 100$ d (business, not calendar, days), so a truncated Lévy distribution of $r_1(t)$ is in principle not ruled out for $\tau = 1$ d [13, 28]. Also not ruled out is the distribution suggested in ref. [8]

Our most important finding is that the fat tails remain intact for the SSRM, showing that the fat tails can persist for $\tau > 1$ d when the data lack sign correlations but have long-range correlated absolute values. This finding proves that whatever the choice of the distribution $p(r_1)$ of daily returns, long-range correlations in the volatility are necessary to explain the behavior of $p(r_\tau)$.

An important consequence of these findings is that great care must be taken when trying to study the distributions $p(r_\tau)$ independently of the correlations. Our findings show that this is true for $\tau > 1$ d, and it is possible that a study of higher frequency data, with many daily data points, would show similar behavior for $\tau < 1$ d. The lower renormalization cutoff could conceivably be as small as the resolution of the data set, even as small as 10 s for a high volume American stock.

We also find that the central bell curve region of the distribution of returns is more similar to that of the SDRM than to the SSRM for $\tau > 1$, showing that in this region the real data are more similar to a Gaussian and that Markovian sign correlations in the returns are important in maintaining the Gaussian-like appearance. Finally, we also find that the behavior of the distribution $p(r_\tau)$ is remarkably similar for different τ .

The new results reported here are of broad interest and scientifically important because long-range correlations and fat tailed distributions can be found in many physical, chemical, and biological phenomena. Moreover, the prices of many financial derivative products depend only on the distribution of returns. The existence of long-range volatility correlations and fat tailed distributions underlying financial time series has been known for some time. What was not fully understood is the origin of the fat tails—which turns out to persist for large lags mainly because of long-range correlations in the volatility. Note that exponentially decaying (i.e., not power law) correlations cannot lead on their own to fat tails at large lags. A systematic study of the S&P 500 index by Gopikrishnan *et al.* [10] had found that the observed scaling of the distributions is due to time dependencies. Here, we have shown that shuffling only the signs, but not the absolute values, of the returns allows the fat tails to persist—hence the fat tails are due to volatility correlations rather than any other kind of time dependency.

We comment on the finding that the real data are more similar to the SDRM in the central bell curve region, but more similar to the SSRM in the fat tailed region. Fig. 3 shows that the absence of volatility correlations causes a collapse of the fat tails in the SDRM, but that the absence of sign correlations causes large deviations in the central bell curve region in the SSRM, for $\tau > 1$. The real data appear qualitatively somewhere in between the SDRM and the SSRM. The implication is that the error of neglecting sign correlations can somehow “compensate” the error of ignoring long-range volatility correlations for small price changes. As a result, for $\tau > 1$ the central region is deceptively well described by a Gaussian. The two errors cancel each other out, hence the success of Bachelier’s century-old Gaussian theory of stock returns. The Gaussian assumption, known to be only approximate, was nevertheless the basis of the Economics Nobel Prize in 1997 (awarded for the work that led to the Black-Scholes Options Pricing Theory). The main flaw in the Gaussian theory is that it cannot explain the fat tails that point to relatively rare but large events, such as the great correction of 1987. There are other known phenomena in which the cancellation of two errors has led to surprisingly good models, a classic example being the original theory of polymer melts by Flory [36], in which the error in estimating repulsive and attractive energies could effectively “cancel” each other out, such that the model became better than otherwise would be expected.

In summary, our findings indicate that the fat tailed distributions of stock returns are mainly due to volatility correlations. More generally, we have shown that fat tailed distributions can arise from long-range correlations in the absolute value of any time series. We note that the shuffling techniques we have developed here are general and can be applied to the study of time series generated by other dynamically rich complex systems that present similar challenges.

We thank the Brazilian agencies CNPq (for support) and CAPES (for funding the visit to Brazil of MS) and P. Gopikrishnan, P. Ch. Ivanov and H. Eugene Stanley for very helpful comments.

-
- [1] R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, 2000).

- [2] E. W. Montroll and W. W. Badger, *Introduction to Quantitative Aspects of Social Phenomena* (Gordon and Breach, New York, 1974).
- [3] M. F. Shlesinger, G. Zaslavsky and U. Frisch, eds., *Lévy Flights and Related Topics in Physics* (Springer, Berlin, 1995).
- [4] J.-P. Bouchaud, A. Maticz and M. Potters, Phys. Rev. Lett. **87**, 228701 (2001).
- [5] V. M. Eguiluz and M. G. Zimmermann, Phys. Rev. Lett. **85**, 5659 (2000).
- [6] B. Podobnik, P. Ch. Ivanov, Y. Lee, A. Chessa and H. E. Stanley, Europhys. Lett. **50(6)**, 711 (2000).
- [7] B. Podobnik, P. Ch. Ivanov, Y. Lee and H. E. Stanley, Europhys. Lett. **52**, 491 (2000).
- [8] M. Pasquini and M. Serva, Euro. Phys. J. B **16**, 195 (2000).
- [9] Y. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C.-K. Peng and H. E. Stanley, Phys. Rev. E. **60**, 1390 (1999).
- [10] P. Gopikrishnan, V. Plerou, L. A. N. Amaral, M. Meyer, and H. E. Stanley, Phys. Rev. E **60**, 5305 (1999).
- [11] M. Pasquini and M. Serva, Econ. Lett., **65**, 275 (1999).
- [12] R. N. Mantegna and H. E. Stanley, Nature **383**, 587 (1996).
- [13] R. N. Mantegna and H. E. Stanley, Nature **376**, 46 (1995).
- [14] B. B. Mandelbrot, J. Bus. **36**, 394 (1963).
- [15] L. P. Kadanoff, Simulation **16**, 261 (1971).
- [16] R. N. Mantegna, Physica A **179**, 232 (1991).
- [17] W. Li, Int. J. Bifurcation and Chaos **1**, 583 (1991).
- [18] H. Takayasu, H. Miura, T. Hirabayashi and K. Hamada, Physica A **184**, 127 (1992).
- [19] P. Bak, K. Chen, J. A. Scheinkman and M Woodford, Ricerche Economiche **47**, 3 (1993).
- [20] J.-P. Bouchaud and D. Sornette, J. Phys. I France **4**, 863 (1994).
- [21] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner and Y. Dodge, Nature **381**, 767 (1996).
- [22] M. Mevy and S. Solomon, Int. J. Mod. Phys. C **7**, 595 (1996).
- [23] S. Galluccio, G. Caldarelli, M. Marsili and Y.-C. Zhang, Physica A **245**, 423 (1997).
- [24] P. Bak, M. Paczuski and M. Shubik, Physica A **246**, 430 (1997).
- [25] N. Vandewalle and M. Ausloos, Physica A **246**, 454 (1997).
- [26] Y. Liu, P. Cizeau, M. Meyer, C.-K. Peng and H. E. Stanley, Physica A **245**, 437 (1997).
- [27] P. Cizeau, Y. Liu, M. Meyer, C.-K. Peng and H. E. Stanley, Physica A **245**, 441 (1997).

- [28] B. Podobnik, K. Matia, A. Chessa, P. Ch. Ivanov, Y. Lee and H. E. Stanley, *Physica A* **300**, 300 (2001).
- [29] J. E. Ingersoll, *Theory of Financial Decision Making* (Rowman & Littlefield, Savage, 1987).
- [30] L. Bachelier, *Ann. Sci. École Norm. Sup.* **17**, 21 (1900).
- [31] R. N. Mantegna and H. E. Stanley, *Phys. Rev. Lett.* **73**, 2946 (1994).
- [32] M. Greene and B. Fielitz, *J. Financial Econ.* **4**, 339 (1977).
- [33] V. Akgiray, *J. Bus.* **62**, 55 (1989).
- [34] A. W. Lo, *Econometrica* **59**, 1279 (1991).
- [35] N. Crato and P. Rothman, *Econ. Lett.* **45**, 287 (1994).
- [36] P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, New York, 1953).

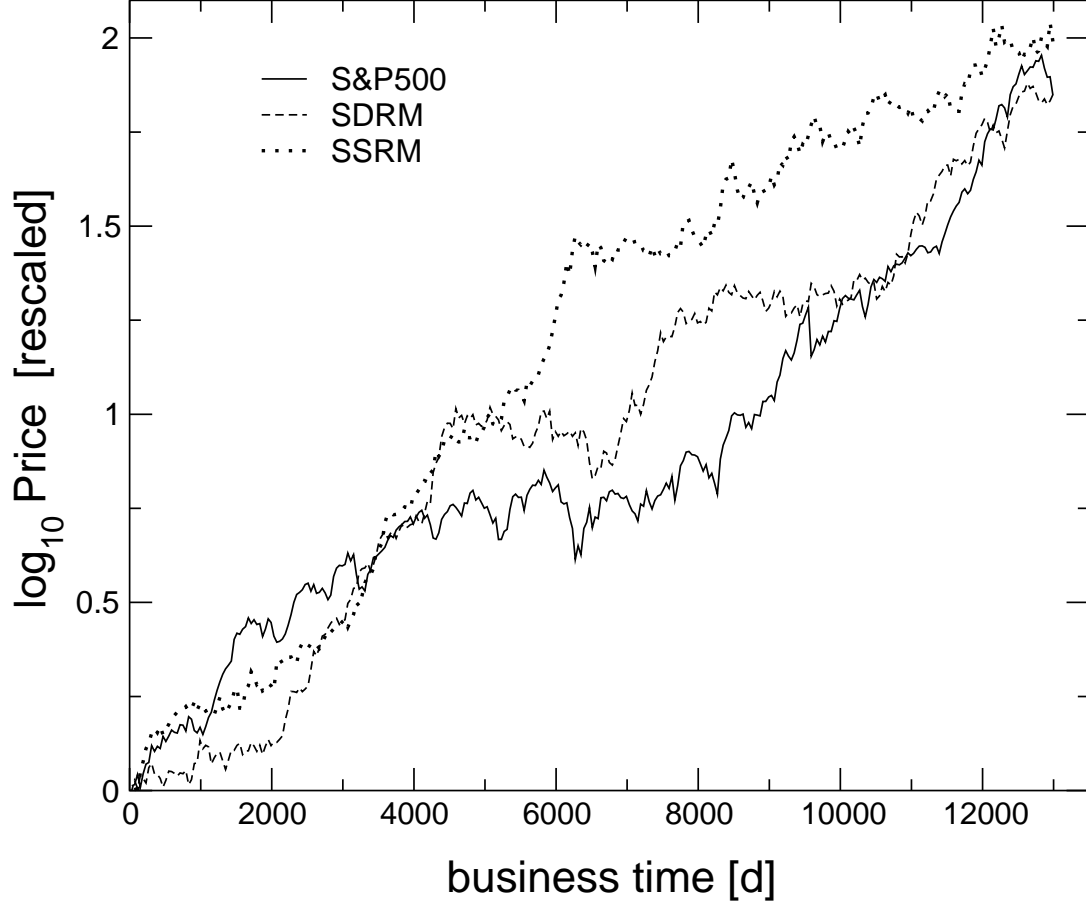


FIG. 1: S&P 500 index, shown on a base 10 logarithmic scale, offset to zero. Also shown are SDRM which has completely uncorrelated returns but an identical $\tau = 1$ d distribution, and SSRM, which has an identical probability distribution of the absolute daily returns, but lacks sign correlations.

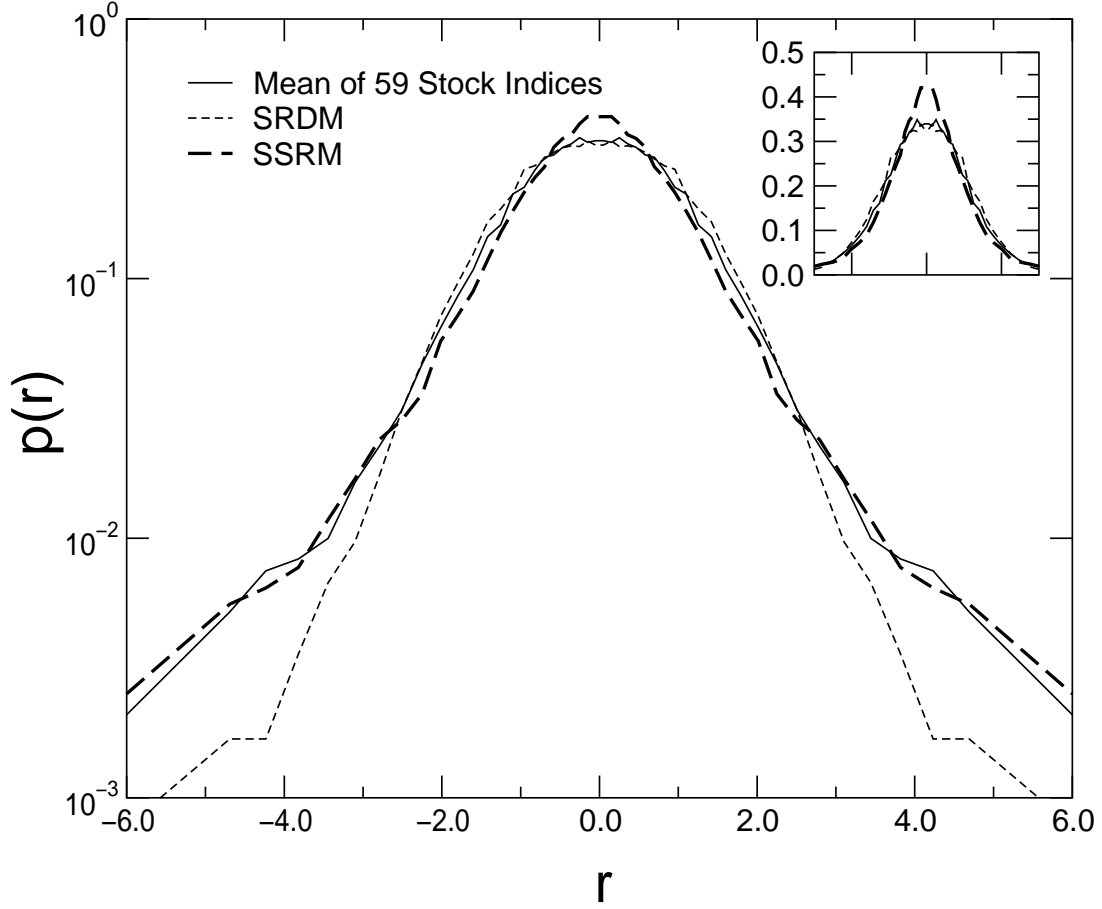


FIG. 2: Symmetrized probability density distribution $p(r)$ of the returns r_τ measured over periods $\tau = 10$ d for 59 stock indices. Also shown are the SSRM and SDRM for $\tau = 10$ d. These distributions are typical of $\tau > 1$ d. We find there is a fat tail in SSRM but not in SDRM, indicating that the origin of the fat tails lies in known long-range correlations in the absolute returns. Inset follows a linear (not semilog) scale. The distribution of r_1 has been normalized to unit variance. For $\tau = 1$ d all three distributions are identical (not shown).

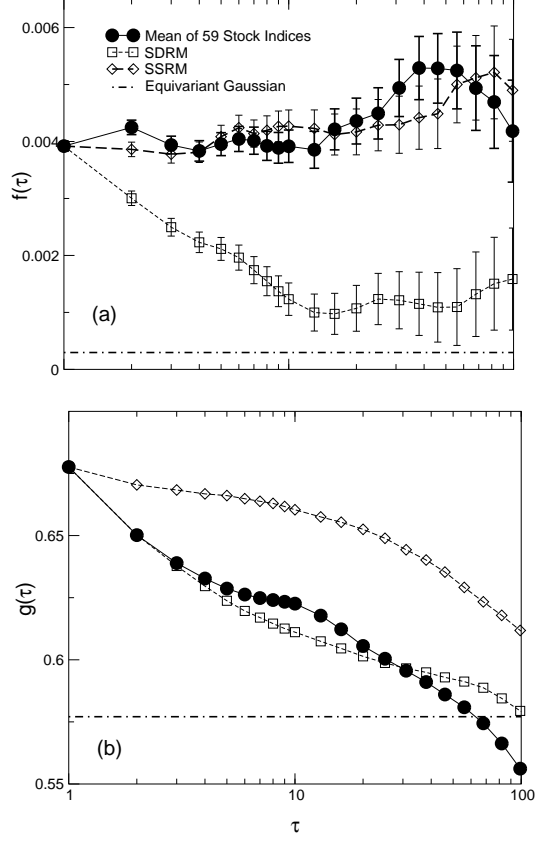


FIG. 3: (a) Mean characteristic function $f(\tau)$ for 59 stock indices, along with the SSRM and SDRM controls, focusing on the fat tails. The values for the SDRM become significantly lower with τ , indicating that the tails are less fat for the shuffled data. The values for the SSRM, however, are remarkably consistent with the original data, showing that known long-range volatility correlations are the real cause of the observed non-Gaussian fat tailed distributions. The loss of the fat tail for the SDRM thus rules out a true Lévy stable distribution. (b) Mean characteristic function $g(\tau)$ for the same datasets. Note that the SSRM has many more returns near zero for $\tau > 1$ d, leading to a higher value of g . This result shows that sign correlations in the real data play an important role that counteract the volatility correlations. Another result seen in (a) and (b) is that an equivariant Gaussian approximation is extremely good for small r (as seen from $\Delta g/g \simeq 20\%$), but very bad for large r (since $\Delta f/f \simeq 1000\%$), a finding potentially important for options pricing theory.